Path coupling without contraction

Magnus Bordewich [a,*], Martin Dyer [b]

[a] *Department of Computer Science, Durham University, Durham, UK.*

[b] *School of Computing, University of Leeds, Leeds LS2 9JT, UK.*

Abstract

Path coupling is a useful technique for simplifying the analysis of a coupling of a Markov chain. Rather than defining and analysing the coupling on every pair in $\Omega \times \Omega$, where Ω is the state space of the Markov chain, analysis is done on a smaller set $S \subseteq \Omega \times \Omega$. If the coefficient of contraction β is strictly less than one, no further analysis is needed in order to show rapid mixing. However, if $\beta = 1$ then analysis (of the variance) is still required for all pairs in $\Omega \times \Omega$. In this paper we present a new approach which shows rapid mixing in the case $\beta = 1$ with a further condition which only needs to be checked for pairs in S, greatly simplifying the work involved. We also present a technique applicable when $\beta = 1$ and our condition is not met.

Key words: Markov chain, Markov chain Monte Carlo, Path coupling, coupling

1 Introduction

A *coupling* for a Markov chain is simply a joint evolution of two copies of the chain. See, for example, Mitzenmacher and Upfal [15]. The origin of the coupling technique for the analysis of Markov chains dates back to 1938 (Doeblin [5]). However, its application to the quantitative analysis of mixing rates did not occur until 1983 (Aldous [1]). Its use for the approximation of hard counting problems began only in 1995 (Jerrum [12]). Nonetheless, during the last decade, coupling has become a standard technique for proving rapid mixing of Markov chains.

The usual approach is to define an appropriate metric on the state space and to show that each step of a suitable coupling produces a contraction in distance

* Corresponding author.

Email addresses: m.j.r.bordewich@durham.ac.uk (Magnus Bordewich), dyer@comp.leeds.ac.uk (Martin Dyer).

between any pair of states. The difficulty with applying this method, however, is that good couplings may not be easy to find, and can be difficult to analyse directly. The evolution of the coupling must be defined and analysed for all pairs of states in the chain. *Path coupling* [3] was developed as a tool to help in the construction and analysis of couplings. It allows the task to be restricted to a smaller set of pairs of states.

When true contraction occurs, the method is easy to apply, at least in principle. However, there is a version of the technique which applies when we can only show *non-expansion*, so that no pair undergoes an increase in expected distance. Here we must show additionally that there is some variance in the distance at each step. However, we are again obliged to show this for *all* pairs, not simply for the smaller set on which we have shown non-expansion. Moreover, a crude lower bound on the variance will give a poor bound on mixing time. Establishing a good bound on variance can be a difficult, or at best tedious, task. In the worst case this work may be unavoidable, since it may even be true that rapid mixing does not occur. However, we will show that general principles can be applied in most situations to deduce a good lower bound on the variance.

We prove a general theorem which can be applied directly in most cases. Where this theorem does not apply, we give a modification of this idea which will often be applicable. We illustrate our methods with examples taken from the literature [2,9,11,16] which have made use of the non-expansion case of path coupling. In all these examples we improve the bound on mixing time established in the source paper. Some of these papers originally used lengthy arguments to bound the variance. We can replace these by a more straightforward, and relatively concise, analysis.

We hope that the techniques we develop here can make the non-expansive case of path coupling as routine to apply as the contractive. In the next section we give a brief review of coupling and path coupling, and state a version of our main result. In Section 3 we give an illustrative example of the approach. Section 4 gives the proof of the main theorem and an example of its use. Section 5 gives an additional approach to proving rapid mixing in situations when our main theorem is not directly applicable. This deals with the case in which the state space of the Markov chain contains transient states, and there actually may be no variance in the distance for some pairs of states of the chain.

2 Background and results

2.1 A review of path coupling

Let Ω be a finite set and let $\mathcal{M}$ be a Markov chain with state space Ω, transition matrix P and unique stationary distribution π. In order for a Markov chain to be useful for almost uniform sampling or approximate counting, it must converge quickly towards its stationary distribution π. We make this notion more precise below. If the initial state of the Markov chain is x then the distribution of the chain at time t is given by $P_x^t(y) = P^t(x,y)$. The *total variation distance* of the Markov chain from π at time t with initial state x, is defined by

$$\mathrm{d_{TV}}(P_x^t, \pi) = \frac{1}{2}\sum_{y\in\Omega} |P^t(x,y) - \pi(y)|.$$

Let $\tau_x(\varepsilon)$ denote the least value T such that $\mathrm{d_{TV}}(P_x^t, \pi) \le \varepsilon$ for all $t \ge T$. The *mixing time* of $\mathcal{M}$, denoted by $\tau(\varepsilon)$, is defined by $\tau(\varepsilon) = \max\{\tau_x(\varepsilon) : x \in \Omega\}$. A Markov chain is said to be *rapidly mixing* if the mixing time is bounded above by some polynomial in n and $\log(\varepsilon^{-1})$, where n is a measure of the size of the elements in Ω.

There are relatively few methods for proving that a Markov chain is rapidly mixing. One such method is *coupling*. A coupling for $\mathcal{M}$ is a stochastic process (X_t, Y_t) on $\Omega\times\Omega$ such that each of (X_t) and (Y_t), considered marginally, is a faithful copy of $\mathcal{M}$. The *Coupling Lemma* [1] states that

$$\mathrm{d_{TV}}(P_x^t, \pi) \le \Pr[X_t \ne Y_t],$$

i.e. the total variation distance for $\mathcal{M}$ at time t is bounded by the probability that the process has not coupled. If we take a metric d on $\Omega\times\Omega$, and let

$$\begin{aligned}
\beta(\mathcal{M},\mathcal{C}) &= \max_{(x,y)\in\Omega^2} \mathbb{E}_{\mathcal{C}}\Big[\mathrm{d}(X_{t+1},Y_{t+1})/\mathrm{d}(x,y) \mid X_t = x, Y_t = y\Big],\\
\sigma^2(\mathcal{M},\mathcal{C}) &= \min_{(x,y)\in\Omega^2} \mathbb{E}_{\mathcal{C}}\Big[\Big(\mathrm{d}(X_{t+1},Y_{t+1}) - \mathrm{d}(x,y)\Big)^2 \mid X_t = x, Y_t = y\Big].
\end{aligned}$$

We will simply write β and σ^2 if the Markov chain and coupling are obvious from the context. It follows from the Coupling Lemma (see for example [6, Theorem 2.1] that if

(i) $\beta(\mathcal{M},\mathcal{C}) < 1$, or
(ii) $\beta(\mathcal{M},\mathcal{C}) = 1$ and $\sigma^2(\mathcal{M},\mathcal{C}) > 0$,

then the chain $\mathcal{M}$ is ergodic and its mixing time can be bounded.

A very useful extension of coupling is the *path coupling* method [3]. In this,

one need only define a coupling on a subset S of $\Omega \times \Omega$. Let $\mathcal{G} = (\Omega, S)$ be the corresponding digraph. Relative to S, a *path coupling* $\mathcal{P}$ for $\mathcal{M}$ is specified by giving only the distributions

$$\Pr(X_{t+1} = x', Y_{t+1} = y' \mid X_t = x, Y_t = y) \qquad (x, y) \in S.$$

The coupling $\mathcal{P}$ is then formed by composing these distributions along paths in $\mathcal{G}$. See [7], for example, for details. Thus, for each pair $(x, y) \in \Omega^2$, we specify a path $x = z_0, z_1, \ldots, z_r = y$ and compose the couplings for (z_{i-1}, z_i) $(i = 1, \ldots, r)$ to give the coupling for x, y.

The path coupling method requires d to be a *path metric* for $\mathcal{G}$, i.e. for all $(x, y) \in \Omega^2$, there is a path $x = z_0, z_1, \ldots, z_r = y$ in $\mathcal{G}$ such that

$$\mathrm{d}(x, y) = \sum_{i=1}^{r} \mathrm{d}(z_{i-1}, z_i).$$

The triangle inequality implies that such a path must be a *shortest path* in $\mathcal{G}$ under the edge weighting given by d. Therefore, we will call such a path a *geodesic*. The paths used to construct the path coupling $\mathcal{P}$ are then chosen to be geodesics. Note that $\mathcal{P}$ may not be unique if geodesics are not unique in $\mathcal{G}$.

A judicious choice of S can greatly simplify the proof of rapid mixing of a Markov chain by coupling. The pairs in S (i.e. edges of $\mathcal{G}$) need not be transitions of $\mathcal{M}$, and vice versa. However, an important special case is where $\mathcal{G}$ is the underlying graph of $\mathcal{M}$ induced by transitions with positive probability and d is edge distance in $\mathcal{G}$. Then d is called *transition distance*.

For a path coupling $\mathcal{P}$ with respect to $S \subseteq \Omega^2$, we define

$$\hat{\beta}(\mathcal{M}, \mathcal{P}) \;=\; \max_{(x,y)\in S} \mathbb{E}_{\mathcal{P}}[\mathrm{d}(X_{t+1}, Y_{t+1})/\mathrm{d}(x, y) \mid X_t = x, Y_t = y].$$

The path coupling method is then given by the following theorem [6].

Theorem 1. *Let* d *be an integer valued metric defined on* $\Omega \times \Omega$ *which takes values in* $\{0, \ldots, D\}$. *Let* S *be a subset of* $\Omega \times \Omega$ *such that for all* $(X_t, Y_t) \in \Omega \times \Omega$ *there exists a path* $X_t = Z_0, Z_1, \ldots, Z_r = Y_t$ *between* X_t *and* Y_t *such that* $(Z_l, Z_{l+1}) \in S$ *for* $0 \le l < r$ *and* $\sum_{l=0}^{r-1} \mathrm{d}(Z_l, Z_{l+1}) = \mathrm{d}(X_t, Y_t)$.

Define a coupling $\mathcal{P} : (X, Y) \mapsto (X', Y')$ *of the Markov chain* $\mathcal{M}$ *on all pairs* $(X, Y) \in S$. *Suppose there exists* $\hat{\beta} \le 1$ *such that* $\mathbb{E}[\mathrm{d}(X', Y')] \le \hat{\beta}\,\mathrm{d}(X, Y)$ *for all* $(X, Y) \in S$.

(i) If $\hat{\beta} < 1$ *then the mixing time* $\tau(\varepsilon)$ *of* $\mathcal{M}$ *satisfies* $\tau(\varepsilon) \le \dfrac{\log(D\varepsilon^{-1})}{1 - \hat{\beta}}$.

(ii) *If* $\hat{\beta} = 1$ *and* $\Pr[\,\mathrm{d}(X_{t+1}, Y_{t+1}) \neq \mathrm{d}(X_t, Y_t)] \geq \alpha$ *for some* $\alpha > 0$ *and all* t *then*

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\alpha} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

Note that, for cases in which d is an integer valued metric, the condition $\Pr[\,\mathrm{d}(X_{t+1}, Y_{t+1}) \neq \mathrm{d}(X_t, Y_t)] \geq \alpha > 0$ is equivalent to the previously stated condition $\sigma^2(\mathcal{M}, \mathcal{P}) > 0$.

The key fact used in path coupling is that we always have $\beta(\mathcal{M}, \mathcal{P}) = \hat{\beta}(\mathcal{M}, \mathcal{P})$ [6, Theorem 2.2]. This greatly simplifies the estimation of $\beta(\mathcal{M}, \mathcal{P})$. However, if $\beta(\mathcal{M}, \mathcal{P}) = 1$, path coupling does not assist in the estimation of $\sigma^2(\mathcal{M}, \mathcal{P})$. Of course, precise estimation is less crucial than for $\beta(\mathcal{M}, \mathcal{P})$, and a crude lower bound will usually suffice. Nevertheless, even placing such a lower bound on $\sigma^2(\mathcal{M}, \mathcal{P})$ is not always straightforward, since the way the coupling is transmitted along paths can be far from obvious.

2.2 Our results

The main result of this paper is that even when $\beta(\mathcal{M}, \mathcal{P}) = 1$, if for all pairs (v, w) in S there is a positive probability that one step of the chain $\mathcal{M}$ from state v moves "directly towards" w, then $\mathcal{M}$ mixes rapidly. Section 4 contains a formal definition of this condition and statement of the result. This is a condition only on pairs in S and it is met in many natural situations. In particular, in the common situation in which the metric on Ω is transition distance, then our result is as follows.

Theorem 2. *Suppose we have a path coupling* $\mathcal{P}$ *for a Markov chain* $\mathcal{M}$ *with respect to the transition distance* d*, where* $S \subseteq \Omega^2$ *is the set of pairs of states at distance 1, and suppose* $D = \max\{\,\mathrm{d}(x, y) : x, y \in \Omega\}$*. Let* P *be the transition matrix for* $\mathcal{M}$*, and let* p *be the minimum non-zero transition probability. Let* $\mathcal{M}^*$ *be the Markov chain with transition matrix* $P' = (P + pI)/(1 + p)$*. If* $\beta(\mathcal{M}, \mathcal{P}) \leq 1$ *then*

(i) *the mixing time of* $\mathcal{M}^*$ *is bounded above by* $\tau_1(\varepsilon) = \lceil p^{-1} e D^2 \rceil \lceil \log \varepsilon^{-1} \rceil$*,*
(ii) *at the random time* $\tau_2(\varepsilon) = \mathrm{Bin}(\tau_1(\varepsilon), (1 + p)^{-1})$*, where* Bin *denotes a binomially distributed random variable,* $\mathcal{M}$ *is within* ε *of the stationary distribution.*

Theorem 2 is a direct corollary of the more general Theorem 7 below. Note the similarity between the time $\tau_1(\varepsilon)$ here and $\tau(\varepsilon)$ in Theorem 1. In essence, we define a new Markov chain $\mathcal{M}^*$ to which we can apply Theorem 1. Moreover $\mathcal{M}^*$ is simply $\mathcal{M}$ with self loop probabilities increased by $\frac{p}{1+p}$.

3 An illustrative example

We begin with an example which highlights the essential construction of our main theorem and demonstrates the necessity of a condition such as we require. Consider the undirected cycle C_n, so $\Omega = V(C_n)$. Our Markov chain $\mathcal{M}$ will be a simple random walk on this graph. If $X_t = v$, then X_{t+1} is given by choosing either neighbour of v with probability $\frac{1}{2}$. Now $\mathcal{M}$ is ergodic if and only if n is odd. Define S to be the set of edges of C_n, and d to be the shortest path distance in C_n. Let $(X_t, Y_t) = (v, w) \in S$, and let v' be the neighbour of v that is not w, and w' the neighbour of w that is not v. We wish to construct a path coupling with $\hat{\beta} \leq 1$ in the following situation.

Distance

	v	w'
w	1	1
v'	1	3

Coupling

	v	w'	
w	?	?	$\frac{1}{2}$
v'	?	?	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	1

Clearly, the only possible path coupling $\mathcal{P}$ with $\hat{\beta} \leq 1$ is that in which both copies of the chain move in the same direction around the cycle, i.e.

$$(X_{t+1}, Y_{t+1}) = \begin{cases} (w, w') & \text{with probability } \frac{1}{2}, \\ (v', v) & \text{with probability } \frac{1}{2}. \end{cases}$$

Then $\hat{\beta} = 1$, but it is clear that X and Y will never couple, whether or not $\mathcal{M}$ mixes.

It follows that some additional condition is required in order to ensure coupling. The condition we use here is that there is some probability that X_{t+1} has moved along the path from X_t towards Y_t. Then we arrange things so that it is does not matter what value Y_{t+1} takes in the coupling.

Returning to our example, we see that this condition is satisfied, since there is probability $\frac{1}{2}$ that the step X_t to X_{t+1} is a move in the direction of Y_t, even though Y_{t+1} unhelpfully moves away. Suppose we add a small probability that the state remains unchanged at each step of the chain, and we introduce a global coupling in which there is a probability that $Y_t = Y_{t+1}$ while X_{t+1} moves along the path from X_t to Y_t. Under these circumstances we will have introduced some variance, and will be able to bound the mixing time. We form a new Markov chain $\mathcal{M}^*$ in which X_t remains unchanged with probability $\frac{1}{3}$, and moves to each neighbour with probability $\frac{1}{3}$, and a new global coupling $\mathcal{C}$ as follows. Let $(X_t, Y_t) \in \Omega^2$, and suppose the shortest path from X_t to Y_t is $X_t = v_0, v_1, \ldots, v_{d-1}, v_d = Y_t$. Let v_{-1} be the neighbour of v_0 that is not v_1,

and v_{d+1} be the neighbour of v_d that is not v_{d-1}.

$$(X_{t+1}, Y_{t+1}) = \begin{cases} (v_1, v_d) & \text{with probability } \frac{1}{3}, \\ (v_0, v_{d+1}) & \text{with probability } \frac{1}{3}, \\ (v_{-1}, v_{d-1}) & \text{with probability } \frac{1}{3}. \end{cases}$$

Notice that the marginal distributions are in agreement with $\mathcal{M}^*$, and that

$$\beta(\mathcal{M}^*, \mathcal{C}) = \left(\tfrac{1}{3}(d-1) + \tfrac{1}{3}(d+1) + \tfrac{1}{3}d\right)/d = 1,$$

but we have now introduced some variance in $\mathrm{d}(X_t, Y_t)$. A martingale argument, as used in [8,14], now implies a bound on the mixing time of $\mathcal{M}^*$. Since $\mathcal{M}^*$ is simply $\mathcal{M}$ slowed down by a factor of 3/2, we can deduce that X_T will be approximately uniform at a suitable randomized stopping time T, even in the case where $\mathcal{M}$ does not actually mix as a Markov chain.

4 Main Theorem

For a more general case than the above example, for any $\delta > 0$ we define

$$p(\delta) = \min_{(v,w)\in S} \Pr(\mathrm{d}(v, X_{t+1}) \geq \delta, \mathrm{d}(v, X_{t+1}) + \mathrm{d}(X_{t+1}, w) = \mathrm{d}(v, w) \mid X_t = v).$$

That is, $p(\delta)$ is the minimum probability over all $(v, w) \in S$ that X_{t+1} moves a distance at least δ from v along the geodesic to w. For any fixed $\delta > 0$, with $p = p(\delta)$, let $\mathcal{M}^*$ be the Markov chain with transition probabilities

$$\begin{aligned} \Pr_{\mathcal{M}^*}(X_{t+1} = x' \mid X_t = x) &= \frac{\Pr_{\mathcal{M}}(X_{t+1} = x \mid X_t = x) + p}{1+p} && (x' = x) \\ &= \Pr_{\mathcal{M}}(X_{t+1} = x' \mid X_t = x)/(1+p) && \text{otherwise.} \end{aligned}$$

Clearly $\mathcal{M}^*$ has the same equilibrium distribution as $\mathcal{M}$. Our first lemma shows that if we have $p(\delta) > 0$ then there is a coupling for $\mathcal{M}^*$ which has non-expansion and non-zero variance.

Remark 3. If the metric used is the transition distance of the Markov chain, and S is the set of pairs of states at transition distance 1, then $p(1)$ is the smallest non-zero transition probability.

Remark 4. The general setting, in which there is some $\delta > 0$ for which $p(\delta) > 0$, may be reduced to the above case as follows. Even if d is not an integer valued metric, we may still rescale without loss of generality, so that the minimum non-zero distance is 1. We may also take a minimal subset $S' \subseteq S$ which gives rise to the same metric. For such a minimal subset, if $(v, w) \in S'$

then there is no $u \in \Omega$ for which $\mathrm{d}(v,u) + \mathrm{d}(u,w) = \mathrm{d}(v,w)$, since in this case we could remove (v,w) from S'. Thus the only possible step on the geodesic from v to w is the transition $v \to w$.

Lemma 5. *If $\beta(\mathcal{M},\mathcal{P}) \le 1$ and $p = p(\delta) > 0$, then the modified Markov chain $\mathcal{M}^*$ has a coupling $\mathcal{C}$ such that $\beta(\mathcal{M}^*,\mathcal{C}) \le 1$ and $\sigma^2(\mathcal{M}^*,\mathcal{C}) \ge p\delta^2/(1+p)$.*

Proof. We construct $\mathcal{C}$ by modifying the path coupling $\mathcal{P}$. The basic idea will be that Y remains fixed, i.e. $Y_{t+1} = Y_t$, whenever X moves directly towards it, i.e. when the event defining $p(\delta)$ holds. The technical problem is to ensure that $\sigma^2(\mathcal{M}^*,\mathcal{C}) > 0$ while retaining $\beta(\mathcal{M}^*,\mathcal{C}) \le 1$. This will mean that $\mathcal{C}$ may no longer be a path coupling with respect to S.

Let $\vartheta = p/(1+p)$. First we extend $\mathcal{P}$ to a path coupling $\mathcal{P}^*$ for $\mathcal{M}^*$ by coupling the new self loop probabilities for each pair of states in S. Clearly $\beta(\mathcal{M}^*,\mathcal{P}^*) = \vartheta + (1-\vartheta)\beta(\mathcal{M},\mathcal{P})$, so $\beta(\mathcal{M}^*,\mathcal{P}^*) \le 1$ if and only if $\beta(\mathcal{M},\mathcal{P}) \le 1$. Consider $(x,y) \in \Omega^2$ with path $x, z_1, \ldots, z_{k-1}, y$. Let us abbreviate (X_{t+1}, Y_{t+1}) to (X,Y), and let the coupling $\mathcal{P}^*$ have transition probabilities

$$P(\xi,\eta) = \Pr_{\mathcal{P}^*}((X,Y) = (\xi,\eta) \mid X_t = x, Y_t = y).$$

Consider the sets

$$\mathcal{E} = \{\, (\xi,\eta) : \mathrm{d}(x,\xi) \ge \delta, \mathrm{d}(x,\xi) + \mathrm{d}(\xi,z_1) = \mathrm{d}(x,z_1), |\mathrm{d}(\xi,\eta) - \mathrm{d}(x,y)| < \delta \,\},$$
$$\mathcal{E}' = \{\, (\xi,\eta) : \mathrm{d}(x,\xi) \ge \delta, \mathrm{d}(x,\xi) + \mathrm{d}(\xi,z_1) = \mathrm{d}(x,z_1)\}.$$

In an abuse of notation we also denote the event that $(X,Y) \in \mathcal{E}$ by $\mathcal{E}$ and the event that $(X,Y) \in \mathcal{E}'$ by $\mathcal{E}'$. Thus $\mathcal{E}'$ is the event that ξ is on the geodesic to z_1 in $\mathcal{P}^*$, and $\mathcal{E} \subseteq \mathcal{E}'$ is the sub-event that the distance between ξ and η is not then appreciably less than $\mathrm{d}(x,y)$.

Note that $\Pr(\mathcal{E}') \ge \vartheta$. Let $\gamma = \Pr(\mathcal{E})$. Then $|\mathrm{d}(X,Y) - \mathrm{d}(x,y)| \ge \delta$ with probability at least

$$\Pr(\mathcal{E}' \setminus \mathcal{E}) = \Pr(\mathcal{E}') - \Pr(\mathcal{E}) \ge \max\{\vartheta - \gamma, 0\}.$$

Thus the event $\mathcal{E}' \setminus \mathcal{E}$ is "good", since it makes a positive contribution to the variance $\sigma^2(\mathcal{M}^*,\mathcal{P}^*)$. However, it may have negligible probability. If so, we can create variance by modifying the coupling, as we will now show. Thus we balance existing variance, associated with the event $\mathcal{E}' \setminus \mathcal{E}$, with variance we introduce, associated with the event $\mathcal{E}$. We will require the following easy facts.

Proposition 6.

(i) $\mathrm{d}(\xi,y) = \mathrm{d}(x,y) - \mathrm{d}(x,\xi)$ *for all* $(\xi,\eta) \in \mathcal{E}$,
(ii) $(x,\eta) \notin \mathcal{E}'$ *(and hence* $(x,\eta) \notin \mathcal{E}$*) for all* $\eta \in \Omega$,

(iii) $(\xi, y) \notin \mathcal{E}$ *for all* $\xi \in \Omega$.

Proof.

(i) $\begin{aligned} \mathrm{d}(\xi, y) &\leq \mathrm{d}(\xi, z_1) + d(z_1, y) = \mathrm{d}(x, z_1) - \mathrm{d}(x, \xi) + \mathrm{d}(z_1, y) \\ &= \mathrm{d}(x, y) - \mathrm{d}(x, \xi) \leq \mathrm{d}(x, \xi) + \mathrm{d}(\xi, y) - \mathrm{d}(x, \xi) = \mathrm{d}(\xi, y). \end{aligned}$

(ii) Otherwise $\delta \leq \mathrm{d}(x, x) = 0 < \delta$.

(iii) Otherwise $\delta \leq \mathrm{d}(x, \xi) = \mathrm{d}(x, y) - \mathrm{d}(\xi, y) \leq |\mathrm{d}(\xi, y) - \mathrm{d}(x, y)| < \delta$. □

Note, however, that $(\xi, y) \in \mathcal{E}'$ is possible. This asymmetry necessitates our use of $\mathcal{E}$, rather than $\mathcal{E}'$, to modify the coupling.

Let $\vartheta' = \min\{\vartheta, \gamma\}$, and for $\gamma > 0$ let $\varphi = \vartheta'/\gamma$. If $\gamma = 0$ let $\varphi = 0$. Define the new global coupling $\mathcal{C}$, based on $\mathcal{P}^*$, to have transition probabilities such that

$$\begin{aligned}
\Pr_{\mathcal{C}}(X = x, Y = y) &= P(x, y) - \vartheta' \\
\Pr_{\mathcal{C}}(X = \xi, Y = y)) &= P(\xi, y) + \varphi \sum_{\eta:(\xi,\eta)\in\mathcal{E}} P(\xi, \eta), && \text{if } \xi \neq x, \\
\Pr_{\mathcal{C}}(X = x, Y = \eta)) &= P(x, \eta) + \varphi \sum_{\xi:(\xi,\eta)\in\mathcal{E}} P(\xi, \eta), && \text{if } \eta \neq y, \\
\Pr_{\mathcal{C}}(X = \xi, Y = \eta) &= (1 - \varphi)P(\xi, \eta), && \text{if } (\xi, \eta) \in \mathcal{E}, \\
\Pr_{\mathcal{C}}(X = \xi, Y = \eta) &= P(\xi, \eta), && \text{otherwise.}
\end{aligned}$$

The effect of this is to transfer probability from $\mathcal{E}$ and the self loop to events in which $X = x, Y \neq y$ or $X \neq x, Y = y$. These events contribute variance.

However, we must first check that the marginal distributions for X and Y are correct so that $\mathcal{C}$ is a valid coupling for $\mathcal{M}^*$. We have, using Proposition 6,

$$\begin{aligned}
\Pr_{\mathcal{C}}(X = x) &= \sum_{\eta} P(x, \eta) - \vartheta' + \varphi \sum_{\eta\neq y:(\xi,\eta)\in\mathcal{E}} P(\xi, \eta) \\
&= \sum_{\eta} P(x, \eta) - \vartheta' + \varphi \sum_{(\xi,\eta)\in\mathcal{E}} P(\xi, \eta) \\
&= \sum_{\eta} P(x, \eta) - \vartheta' + \varphi\gamma \\
&= \sum_{\eta} P(x, \eta) \\
&= \Pr_{\mathcal{P}^*}(X = x).
\end{aligned}$$

For $\xi \neq x$ we have, using Proposition 6 again,

$$\begin{aligned}
\Pr_{\mathcal{C}}(X = \xi) &= P(\xi, y) + \varphi \sum_{\eta:(\xi,\eta)\in\mathcal{E}} P(\xi, \eta) \\
&\qquad + (1-\varphi) \sum_{\eta:(\xi,\eta)\in\mathcal{E}} P(\xi, \eta) + \sum_{\eta\neq y:(\xi,\eta)\notin\mathcal{E}} P(\xi, \eta) \\
&= \sum_{\eta} P(\xi, \eta) \\
&= \Pr_{\mathcal{P}^*}(X = \xi).
\end{aligned}$$

The calculations for Y are similar. Hence $\mathcal{C}$ is a valid coupling for $\mathcal{M}^*$. Let $\mathbb{E}$ and $\mathbb{E}'$ denote expectation under $\mathcal{P}^*$ and $\mathcal{C}$ respectively. It now follows that

$$\begin{aligned}
\mathbb{E}'[\mathrm{d}(X, Y)] &= \mathbb{E}[\mathrm{d}(\xi, \eta)] - \vartheta' \mathrm{d}(x, y) + \varphi\mathbb{E}[\mathrm{d}(\xi, y)\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] \\
&\qquad + \varphi\mathbb{E}[\mathrm{d}(x, \eta)\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] - \varphi\mathbb{E}\Big[\mathrm{d}(\xi, \eta)\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}\Big] \\
&= \mathbb{E}[\mathrm{d}(\xi, \eta)] - \varphi\mathbb{E}[(\mathrm{d}(x, y) - \mathrm{d}(\xi, y) - \mathrm{d}(x, \eta) + \mathrm{d}(\xi, \eta))\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] \\
&= \mathbb{E}[\mathrm{d}(\xi, \eta)] - \varphi\mathbb{E}[(\mathrm{d}(x, \xi) - \mathrm{d}(x, \eta) + \mathrm{d}(\xi, \eta))\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] \\
&\leq \mathbb{E}[\mathrm{d}(\xi, \eta)] \\
&\leq \mathrm{d}(x, y),
\end{aligned}$$

using, respectively, $\vartheta' = \varphi\mathbb{E}[\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}]$, Proposition 6, the triangle inequality and the non-expansion of $\mathcal{P}^*$. Therefore we have $\mathbb{E}'[\mathrm{d}(X, Y)/\mathrm{d}(x, y)] \leq 1$. Using $\max\{\vartheta - \gamma, 0\} = \vartheta - \vartheta'$, we also have

$$\begin{aligned}
\mathbb{E}'[(\mathrm{d}(X, Y) - \mathrm{d}(x, y))^2] &\geq \delta^2 \Pr(\mathcal{E}' \setminus \mathcal{E}) + \varphi\mathbb{E}[(\mathrm{d}(\xi, y) - \mathrm{d}(x, y))^2\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] \\
&= \max\{\vartheta - \gamma, 0\}\,\delta^2 + \varphi\mathbb{E}[\mathrm{d}(x, \xi)^2\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] \\
&= (\vartheta - \vartheta')\delta^2 + \varphi\mathbb{E}[\mathrm{d}(x, \xi)^2\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] \\
&\geq (\vartheta - \vartheta')\delta^2 + \varphi\delta^2\mathbb{E}[\mathbb{1}_{(\xi,\eta)\in\mathcal{E}}] \\
&= (\vartheta - \vartheta')\delta^2 + \vartheta'\delta^2 \\
&= \vartheta\delta^2,
\end{aligned}$$

where we use only the variance contributed by the events $X = \xi \neq x, Y = y$ with the lower bound $\Pr_{\mathcal{C}}(X = \xi, Y = y) \geq \varphi \sum_{\eta:(\xi,\eta)\in\mathcal{E}} P(\xi, \eta)$. Note that there is a trade-off between this and the variance contributed by $\mathcal{E}' \setminus \mathcal{E}$. □

Now we are in a position to show rapid mixing by applying a martingale argument similar to that used in [8,14]. Note that we cannot simply apply the equality case of the path coupling method (Theorem 1), since $\mathcal{C}$ is *not* a path coupling but a global coupling.

Theorem 7. *Let $\mathcal{P}$ be a path coupling for the Markov chain $\mathcal{M}$ relative to metric* d *and $S \subseteq \Omega^2$. If $\beta \leq 1$ and $p = p(\delta) > 0$ for some $\delta > 0$ then*

(i) the modified Markov chain $\mathcal{M}^*$ has mixing time $\tau(\varepsilon) \leq \lceil eD^2/\vartheta\delta^2 \rceil \lceil \log \varepsilon^{-1} \rceil$, where $\vartheta = p/(1+p)$ and $D = \max\{ \mathrm{d}(x,y) : x, y \in \Omega \}$,

(ii) if $\tau^*(\varepsilon)$ denotes the random time $\mathrm{Bin}(\tau(\varepsilon), (1+p)^{-1})$, where Bin denotes the binomial distribution, and $X_{\tau^*(\varepsilon)}$ is the state of the original Markov chain $\mathcal{M}$ at this random time, then the distribution of $X_{\tau^*(\varepsilon)}$ is within ε of the stationary distribution for any starting configuration.

Proof. Let $T_{x,y}$ denote the first time that $X_t = Y_t$ given that $X_0 = x, Y_0 = y$, where (X_t, Y_t) evolve according to the coupling $\mathcal{C}$ defined in Lemma 5. Consider the process $Z_t = Z_t(x,y) = (D - \mathrm{d}(X_t, Y_t))^2 - \vartheta\delta^2 \min\{t, T_{x,y}\}$. Then, by Lemma 5, Z_t is a submartingale since, for $t < T_{x,y}$

$$\begin{aligned}
\mathbb{E}[Z_{t+1}] &= \mathbb{E}[(D - \mathrm{d}(X_t,Y_t) + \mathrm{d}(X_t,Y_t) - \mathrm{d}(X_{t+1},Y_{t+1}))^2] - \vartheta\delta^2(t+1) \\
&= Z_t + 2(D - \mathrm{d}(X_t,Y_t))\mathbb{E}[\mathrm{d}(X_t,Y_t) - \mathrm{d}(X_{t+1},Y_{t+1})] \\
&\qquad + \mathbb{E}[(\mathrm{d}(X_t,Y_t) - \mathrm{d}(X_{t+1},Y_{t+1}))^2] - \vartheta\delta^2 \\
&\geq Z_t + 2(D - \mathrm{d}(X_t,Y_t))\mathrm{d}(X_t,Y_t)(1-\beta) + \vartheta\delta^2 - \vartheta\delta^2 \\
&\geq Z_t,
\end{aligned}$$

and for $t \geq T_{x,y}$ we have $Z_{t+1} = Z_t$. Note also that $T_{x,y}$ is a stopping time, and so we may apply the Optional Stopping Theorem. Hence $\mathbb{E}[Z_{T_{x,y}}] \geq \mathbb{E}[Z_0]$, which implies $D^2 - \vartheta\delta^2\mathbb{E}[T_{x,y}] \geq 0$, and hence $\mathbb{E}[T_{x,y}] \leq D^2/\vartheta\delta^2$. Let $\tau = \lceil eD^2/\vartheta\delta^2 \rceil \lceil \ln \varepsilon^{-1} \rceil$. Considering τ as $\lceil \ln \varepsilon^{-1} \rceil$ independent periods of $\lceil eD^2/\vartheta\delta^2 \rceil$, by Markov's inequality we obtain $\Pr[T_{x,y} > \tau] \leq \varepsilon$. The coupling inequality [15] now gives the first part.

By the definition of $\mathcal{M}^*$, for any initial state X_0 the distribution of X_{τ^*} in $\mathcal{M}$ is exactly the distribution of X_τ in $\mathcal{M}^*$, and is therefore equally close to their common stationary distribution. □

By the second part of Theorem 7, for all practical purposes we may regard the original chain $\mathcal{M}$ as mixing in the same time as $\mathcal{M}^*$. See [13, Chapter 5, Remarks 5.5] for a further discussion of this issue.

Theorem 2 follows directly from Theorem 7.

4.1 Application: Sampling 3-colourings in $\mathbb{Z}^2$

In this section we will present an application which demonstrate the use of Theorem 7. We present a much simpler proof of mixing of Glauber dynamics for 3-colourings in $\mathbb{Z}^2$.

A special case of sampling colourings that has attracted interest in statistical physics is sampling colourings of the grid. In particular we will focus on the

mixing rate of Glauber dynamics $\mathcal{M}_{Gl}$ for sampling 3-colourings of a graph G, where G is an $m \times n$ rectangular section of $\mathbb{Z}^2$ with free boundary, i.e. there are no external restrictions on the colouring of the vertices of degree 2 or 3. The Markov chain $\mathcal{M}_{Gl}$ is given by the following transitions. If the state of $\mathcal{M}_{Gl}$ at time t is X_t, the state at $t+1$ is determined by

(i) selecting a vertex $v \in G$ and a colour $k \in \{1, 2, 3\}$ uniformly at random,
(ii) let X'_t be the colouring obtained by recolouring v colour k
(iii) if X'_t is a proper colouring let $X_{t+1} = X'_t$
otherwise let $X_{t+1} = X_t$.

Goldberg *et al.* [11] showed that $\mathcal{M}_{Gl}$ is rapidly mixing, again by considering an alternative Markov chain $\widetilde{\mathcal{M}}$ with additional moves and then using comparison. The significant results from [11] we shall need are simply that there is a coupling for $\widetilde{\mathcal{M}}$ such that $\beta(\widetilde{\mathcal{M}}) \leq 1$ ([11, Theorem 6.6]), and that the maximum distance between two colourings is $2mn^2$, where S is again all pairs of colourings differing at a single vertex. What the additional moves available are is irrelevant to the current discussion, since the standard moves available in Glauber dynamics are enough to bound $p(1)$. It now remains to observe that $p(1)$ is the probability that the correct vertex and colour are chosen, which in $\widetilde{\mathcal{M}}$ is at least $\frac{1}{4} \cdot \frac{1}{nm} \cdot \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{16nm}$. (See [11] for details.) The following corollary is now immediate from Theorem 7.

Corollary 8. *The Markov chain $\widetilde{\mathcal{M}}^*$, for 3-colourings of an $m \times n$ section of $\mathbb{Z}^2$ has mixing time*

$$\tau(\varepsilon) \leq \lceil e(2mn^2)^2 16nm \rceil \lceil \log \varepsilon^{-1} \rceil \leq 65em^3n^5 \lceil \log \varepsilon^{-1} \rceil.$$

Remark 9. This compares favourably with the mixing time bound for the Markov chain $\widetilde{\mathcal{M}}$ given in [11], of $\tau(\varepsilon) \leq 193em^3n^6 \lceil \log \varepsilon^{-1} \rceil$.

Again, in order to save a factor of $\log \pi^*$, it is necessary to use comparison directly between $\widetilde{\mathcal{M}}^*$ and $\mathcal{M}_{Gl}$. However, as before, since all the transition probabilities in $\widetilde{\mathcal{M}}^*$ differ by a constant factor from those of $\widetilde{\mathcal{M}}$, it is straight forward to adapt the comparison of [11] to yield a mixing time bound for $\mathcal{M}_{Gl}$ of $\tau(\varepsilon) = O(m^4n^8 \log \varepsilon^{-1})$. This may be compared with $\tau(\varepsilon) = O(m^4n^9 \log \varepsilon^{-1})$ given in [11].

Remark 10. It is not so much the moderate improvement in the mixing time that motivates this example, as the illustration of the ease of use of Theorem 7. It is almost immediate from the definition of the chain that we will have rapid mixing in the case of $\beta = 1$, using Theorem 7. In [11], three pages and five cases of detailed argument are required to show that $\sigma^2(\mathcal{M}) > 1/(48mn^2)$.

Remark 11. The Markov chain $\widetilde{\mathcal{M}}$ has probability 3/4 of remaining unchanged at any step. This delay is introduced to force all eigenvalues to be

positive. Since these probabilities of remaining unchanged are all coupled with each other in the analysis of [11], there is enough flexibility for us to apply our adjusted coupling directly to $\widetilde{\mathcal{M}}$. This avoids introducing an additional probability of remaining unchanged, as in Lemma 5. Hence the mixing time bounds given above for $\widetilde{\mathcal{M}}^*$ hold also for $\widetilde{\mathcal{M}}$.

4.2 Application: Sampling Eulerian orientations of the triangular lattice

The results of this paper have recently been used to show rapid mixing of a Markov chain on the set of Eulerian orientations of a triangular subsection G of the triangular lattice [4]. The Markov chain $\mathcal{M}_o$ used is as follows. If the state of $\mathcal{M}_o$ at time t is an Eulerian orientation X_t, then X_{t+1} is obtained by

(i) selecting a bounded face f of G uniformly at random,
(ii) if the three edges bounding f form a directed cycle set $X' = X_t$ *with these edges reversed*, otherwise set $X' = X_t$ (unchanged),
(iii) with probability $1/2$ set $X_{t+1} = X'$, otherwise set $X_{t+1} = X_t$.

As in [11], a Markov chain $\widetilde{\mathcal{M}}_o$ with additional moves is analysed and $\mathcal{M}_o$ is shown to be rapidly mixing by comparison. The additional moves allowed in $\widetilde{\mathcal{M}}_o$ involve reversing the edges bounding several adjacent faces with some probability, in the event that the face selected at random does not form a directed cycle. In [4] it is shown that $\beta(\widetilde{\mathcal{M}}_o) \leq 1$. An appeal to Theorem 2 of this paper completes the proof of rapid mixing, avoiding an involved direct proof bounding $\sigma^2(\widetilde{\mathcal{M}}_o)$. See [4] for full details.

Remark 12. The Markov chain $\mathcal{M}_o$ had previously been analysed in [9]. In this paper a modified Markov chain which may reverse the edges of at most two adjacent faces was introduced. It is claimed in [9] that the latter chain is rapidly mixing, however the proof contains an error.

5 A modified technique when S contains transient states

It seems that Theorem 7 will apply in most situations, since couplings are usually designed to bring the chains into the same state quickly and so usually include transitions which move the two states directly towards each other. However, there is a notable exception. In cases where S contains transient states of the Markov chain, it may be that $p(\delta) = 0$ for all δ. For example, this occurs when considering Glauber dynamics on proper colourings of a graph and the metric is *Hamming distance*, i.e. the number of vertices at which the coupled chains differ. It is then usual to define Ω to be the set of *all* colourings

of the graph, and S to be colourings that differ in exactly one vertex. The Markov chain is then constrained always to recolour vertices properly, so that the stationary distribution has positive probability only on proper colourings. This approach has the advantage of simplifying the analysis, and yielding maximum Hamming distance n between two colourings. Note that defining S to be only the pairs of *proper* colourings which differ at exactly one vertex leads to a maximum distance bounded by Δn.

In this situation, it is even possible that the variance can be zero in a single step, but nevertheless the chain mixes rapidly. This occurs when there are some pairs $(x, y) \in \Omega^2$ such that the distance cannot change after a single step of the coupled chain, but the distance can change after more than one step. Thus we define the r-step variance to be

$$\sigma_r^2(\mathcal{M}, \mathcal{C}) \;=\; \min_{(x,y)\in\Omega^2} \mathbb{E}_{\mathcal{C}}\Big[\Big(\mathrm{d}(X_{t+r}, Y_{t+r}) - \mathrm{d}(x, y)\Big)^2 \mid X_t = x, Y_t = y\Big].$$

If there exists some positive integer r such that $\sigma_r^2 > 0$, then the mixing time of $\mathcal{M}$ can be bounded. This is achieved simply by considering the Markov chain $\mathcal{M}^r$, which is the chain taking r steps of $\mathcal{M}$ at a time. The coupling $\mathcal{C}$ for $\mathcal{M}$ is trivially extended to a coupling $\mathcal{C}^r$ for $\mathcal{M}^r$ and it easily follows that $\beta(\mathcal{M}^r, \mathcal{C}^r) \leq \beta(\mathcal{M}, \mathcal{C})^r$ and $\sigma^2(\mathcal{M}^r, \mathcal{C}^r) = \sigma_r^2(\mathcal{M}, \mathcal{C})$. The standard analysis now yields mixing time bounds for $\mathcal{M}^r$ and therefore $\mathcal{M}$.

Vigoda suggests a similar approach in [16] to claim a mixing time for his chain on graph colourings in the $\beta = 1$ case. We discuss this in Section 5.2 below. Here we bound the r-step variance by extending the idea used to prove Theorem 7. Thus we modify the global coupling over a single step to create a more favourable configuration which permits a change in distance at a subsequent step. However, since it seems difficult to frame a general statement like Theorem 7 in this situation, we will simply give two illustrations. We consider two Markov chains on the set of proper k-colourings in graphs of maximum degree Δ. The first is the well known Glauber dynamics Markov chain $\mathcal{M}_{Gl}$, the second is the "flip" chain of Vigoda [16] $\mathcal{M}_{flip}$.

In both our examples $G = (V, E)$ is a graph of maximum degree Δ, and $\mathcal{K} = \{1, 2, \ldots, k\}$ is the colour set. For $v \in V$, $\Gamma(v)$ will denote the neighbour set of v in G. For any colouring x of G, let $\mathcal{A}_x(v) = \mathcal{K} \setminus \{x(w) : w \in \Gamma(v)\}$. We write $\Omega = \mathcal{K}^V$ for the set of all colourings of G. We will consider colourings $(x, y) \in \Omega^2$ with Hamming distance $\mathrm{d}(x, y)$ between x and y. We will write x_t, y_t for the t-step evolution of x, y under some coupling. The couplings used will be known couplings, or simple modifications of them.

5.1 Sampling 2Δ colourings in graphs

Let $\mathcal{M}_{Gl}$ be the Markov chain given by the Glauber dynamics on proper 2Δ-colourings of G. This is defined analogously to Section 4.1 with colour selection uniformly at random from $\{1, \ldots, 2\Delta\}$, see for example [12]. We know (see, for example, [2]) that this chain is rapidly mixing. However, for the coupling defined next, the coefficient of contraction can be equal to 1, and hence our analysis is relevant. The coupling $\mathcal{C}$ is as follows. Let $(x, y) \in S$ and let w be the unique vertex such that $x(w) \neq y(w)$. If the attempted transition in x is recolouring vertex v with colour c, then the coupled move in y is to attempt to recolour vertex v with colour c' where

$$c' = \begin{cases} x(w) & \text{if } v \in \Gamma(w) \text{ and } c = y(w), \\ y(w) & \text{if } v \in \Gamma(w) \text{ and } c = x(w), \\ c & \text{otherwise.} \end{cases}$$

Standard analysis yields $\beta(\mathcal{M}_{Gl}, \mathcal{C}) \leq 1$ when $k = 2\Delta$.

Suppose G is the complete bipartite graph $K_{n/2,n/2}$ with bipartition V_1, V_2. Note that the $\Delta = n/2$ so we have $k = n$. Let x be the colouring in which the vertices of V_1 are coloured with colours 1 through $n/2$ and the vertices of V_2 are coloured with colours $n/2 + 1$ through n. Let y be the reverse: the vertices of V_1 are coloured $n/2 + 1$ through n and the vertices of V_2 are coloured 1 through $n/2$. It is now clear that if x_1 and y_1 are the one step evolutions of x and y under Glauber dynamics then, *whatever the coupling*, $\mathrm{d}(x_1, y_1)$ remains n. Furthermore, any colouring z such that $\mathrm{d}(x, z) = 1$ and $\mathrm{d}(x, z) + \mathrm{d}(z, y) = \mathrm{d}(x, y)$ must be an improper colouring. Hence we cannot use the approach of Section 3. However, observe that if we choose $v_1 \in V_1$ and $c \in x(V_1 \setminus \{v_1\})$ in the first step, and $v_2 \in V_2$ and colour $x(v_1)$ in the second step, then x_2 and y_2 obtained after two steps are distance $n - 1$ apart (since v_2 is recoloured $x(v_1)$ in both chains).

For a general connected graph, if there is no move available in one step such that the distance changes then we must be in a similar situation. The states x and y must be at distance exactly n and each vertex v must see Δ different colours in $\Gamma(v)$ in x and the remaining Δ colours in y. Thus, in particular, the graph must be Δ-regular. Suppose we modify the global coupling to be the identity coupling for one step. This cannot increase the distance since it is already maximum. Then, if we choose v and a colour $c \neq x(v)$ which is accepted in x, this colour is necessarily rejected in y, yielding x_1 and $y_1 = y$. Then colour c is available for every neighbour of v in both x_1 and y_1 at step two. Hence, for such a pair x and y, the probability of changing the distance in two steps is at least $\left(\frac{n}{n}\frac{\Delta-1}{2\Delta}\right)\left(\frac{\Delta}{n}\frac{1}{2\Delta}\right) = \frac{\Delta-1}{4n\Delta}$. For any other pair x and y there is a step which reduces the distance immediately, and if the second step

does not increase the distance then $\mathrm{d}(x_2, y_2) \neq \mathrm{d}(x, y)$. Since $\beta(\mathcal{M}_{Gl}, \mathcal{C}) \leq 1$, the probability that the distance increases in a single step is at most $\frac{1}{2}$, hence $\sigma_2^2 \geq \frac{1}{4n\Delta}$. This yields a mixing time of $\tau_{\mathcal{M}_{Gl}^2}(\varepsilon) \leq \lceil 4en^3\Delta \rceil \lceil \log \varepsilon^{-1} \rceil$ for $\mathcal{M}_{Gl}^2$ and therefore $\tau_{\mathcal{M}_{Gl}}(\varepsilon) \leq 2\lceil 4en^3\Delta \rceil \lceil \log \varepsilon^{-1} \rceil = O(n^3\Delta \log \varepsilon^{-1})$.

Remark 13. This mixing time can be improved by a constant factor by observing that the variance can only be zero when the distance between the two states is n, and hence the two-step Markov chain needs only be considered at such times.

5.2 Sampling $\frac{11}{6}\Delta$-colourings in graphs

Here $k = \frac{11}{6}\Delta$. Vigoda [16] showed rapid mixing of a chain for sampling k-colourings when $k > \frac{11}{6}\Delta$. This chain swaps the colours of a bi-coloured cluster of vertices at each step. A bi-coloured cluster is defined as follows. From a colouring x, a vertex $v \in G$ and a colour c, the cluster $S_x(v, c)$ is the set of $w \in V$ such that there is an alternating path from v to w using colours c and $x(v)$. Thus $S_x(v, c) = \{v\}$ is equivalent to $c \in \mathcal{A}_x(v)$.

If the state of $\mathcal{M}_{flip}$ at time t is X_t, the state at $t + 1$ is determined by

(i) selecting a vertex $v \in G$ and a colour c uniformly at random,
(ii) let X_t' be the colouring obtained by 'flipping' the cluster $S_{X_t}(v, c)$, that is interchanging colours c and $X_t(v)$ in the cluster,
(iii) let $\alpha = |S_{X_t}(v, c)|$
with probability $\frac{p_\alpha}{\alpha}$ set $X_{t+1} = X_t'$
with probability $1 - \frac{p_\alpha}{\alpha}$ set $X_{t+1} = X_t$.

Since the cluster $S_{X_t}(v, c)$ can be chosen $|S_{X_t}(v, c)|$ ways, it is flipped with probability $\frac{p_\alpha}{nk}$. The values of the constants p_α are determined in [16]; the salient points for us are simply that $p_1 = 1$ and $p_i \leq \frac{13}{46}$ for $i \geq 2$.

In [16] $\mathcal{M}_{flip}$ is proved rapidly mixing via path coupling whenever $k > \frac{11}{6}\Delta$, and hence $\mathcal{M}_{Gl}$ is proved rapidly mixing by comparison in this range. The case of interest here is when $k = \frac{11}{6}\Delta$, i.e. when $\Delta \equiv 0 \pmod 6$. In this situation it is shown in [16] that $\beta(\mathcal{M}_{flip}) \leq 1$.

Vigoda [16] considers the state space of the chain to be the set of all colourings, which is permissible since the chain will never move to an improper colouring from a proper colouring. Since S is taken to be pairs of colourings differing at a single vertex, this reduces the distance between any two states to the number of vertices on which they differ, and hence $D = n$. However, under these circumstances $p(\delta)$ is zero for all $\delta > 0$ since $(X, Y) \in S$ allows Y to be an improper colouring.

Considering $\sigma^2(\mathcal{M}_{flip})$ directly, Vigoda [16] remarks without proof that the bound $\sigma^2 \geq (nk)^{-1-\Delta/6}$ holds. Technically, he cannot guarantee any move that changes the distance in one step, but a change can be achieved in $\frac{\Delta}{6}+1$ steps. This is to be done by recolouring at most $\frac{\Delta}{6}$ vertices without changing the distance, in order to reach a state in which some vertex of difference can be recoloured with a common colour in both chains. However, it is not clear how this recolouring can be done while maintaining the coupling which gives $\beta \leq 1$. Note that we cannot simply switch to (say) the identity coupling to achieve this, as we did in Section 5.1, since this may increase the distance. However, if this recolouring can be done, it implies that $\sigma^2(\mathcal{M}_{flip}^{\Delta/6+1}) > 0$, where this chain makes $\frac{\Delta}{6}+1$ steps of $\mathcal{M}_{flip}$ at each step. This would give a mixing time bound for $\mathcal{M}_{flip}$ of

$$\tau(\varepsilon) \leq \lceil en^2(nk)^{1+\Delta/6}\rceil\lceil\log(\varepsilon^{-1})\rceil = O\Big(k^{1+\Delta/6}n^{3+\Delta/6}\log(\varepsilon^{-1})\Big). \qquad (1)$$

We will now show that under a slightly modified global coupling there is a move that decreases the distance between any two states in at most two steps of the chain. We do this by adjusting the coupling so that if there is no single step which reduces the distance between two colourings, then in one step we release a colour so that in the second step both copies of the chain may recolour some disagreement vertex with this colour.

Let x, y be two copies of the chain coupled using Vigoda's coupling. Let x_1, y_1 and x_2, y_2 be the one and two step evolutions of the coupled chains x and y. For a vertex v and colour $c \in \mathcal{A}_x(v)$ we have $S_x(v,c) = \{v\}$ and with probability at least $1/nk$ chain x chooses cluster $S_x(v,c)$; in which case $x_1(v) = c$ and $x_1(w) = x(w)$ $(w \neq v)$. We do not need to consider all details of Vigoda's coupling, but let us note the facts in the following proposition, whose proof is deferred until later.

Proposition 14. *Let $\kappa = 1 - 2p_2 \geq \frac{16}{42}$. The transmission of the coupling along the path to y can be taken to have the property that*

- *(i) with probability at least $1/nk$ both x and y make a null transition $(x_1 = x, y_1 = y)$;*
- *(ii) if chain x chooses a vertex v and a colour $c \in \mathcal{A}_x(v) \cap \mathcal{A}_y(v)$, then y flips the cluster $S_y(v,c) = \{v\}$;*
- *(iii) if chain x chooses a vertex v and a colour $c \in \mathcal{A}_x(v) \setminus \mathcal{A}_y(v)$ then either*
  - *(a) with probability at least κ chain y flips a cluster $S_y(v,c') = \{v\}$ for some $c' \neq c$; or*
  - *(b) with probability at least κ chain y makes the null transition $(y_1 = y)$.*

Let $v \in V$ be any vertex such that $x(v) \neq y(v)$. Suppose $\mathcal{A}_x(v) \cap \mathcal{A}_y(v) \neq \emptyset$. By (ii), with probability at least $1/nk$ we will have $x_1(v) = y_1(v) = c$ for $c \in \mathcal{A}_x(v) \cap \mathcal{A}_y(v)$, giving $\mathrm{d}(x_1, y_1) = \mathrm{d}(x,y) - 1$. Hence, by (i), with probability at least $(1/nk)^2$ we have $\mathrm{d}(x_2, y_2) \neq \mathrm{d}(x,y)$. We now assume $\mathcal{A}_x(v) \cap \mathcal{A}_y(v) = \emptyset$.

Let s be the number of colours in $\mathcal{A}_y(v)$ which appear exactly once in the multiset $x(\Gamma(v))$, and t the number of colours which appear more than once. Then $s + t \geq k - \Delta$ and $s + 2t \leq \Delta$. Thus $s \geq 2k - 3\Delta$. If $k \geq 11\Delta/6$, then $s \geq 2\Delta/3 \geq 4$ if $\Delta \geq 6$.

Now, let w be any vertex such that $a_x = x(w) \in \mathcal{A}_y(v)$ and a_x appears exactly once in $x(\Gamma(v))$. Let $a_y = y(w)$, and note that $a_x \neq a_y$ since $a_x \in \mathcal{A}_y(v)$, $a_y \notin \mathcal{A}_y(v)$. Thus w is a disagreement between x and y. Let $a \neq a_x$ be such that $a \in \mathcal{A}_x(w)$. This requires $k - \Delta \geq 2$, which is implied by $k \geq 11\Delta/6$ if $\Delta \geq 6$. We will modify the global coupling so that there is some probability that when $x_1(w) = a$ then $y_1(w) \neq a_x$ and thus $a_x \in \mathcal{A}_{x_1}(v) \cap \mathcal{A}_{y_1}(v)$. Using (iii) above, we may restrict our attention to transitions in which x recolours w to a and y recolours only the vertex w in this step.

In case (a), with probability at least κ/nk, x recolours w to colour a, while y recolours w to some other colour a'. If $a' \neq a_x$ then we will leave the coupling unmodified, since we achieve $a_x \in \mathcal{A}_{x_1}(v) \cap \mathcal{A}_{y_1}(v)$. If $a' = a_x$ then $x_1(w) = a$, $y_1(w) = a_x$ with probability at least κ/nk. By (i) we also have that with probability greater than κ/nk, $x_1(w) = a_x$, $y_1(w) = a_y$. Then, on an event with probability $2\kappa/nk$, we modify the coupling so that $y_1(w) = a_y$ when $x_1(w) = a$, and $y_1(w) = a' = a_x$ when $x_1(w) = a_x$. Under this modification we have $a_x \in \mathcal{A}_{x_1}(v) \cap \mathcal{A}_{y_1}(v)$ with probability κ/nk, and incidentally also $x_1(w) = a_x = y_1(w)$ with the same probability.

In case (b), with probability at least κ/nk, x recolours w to colour a, while y makes the null transition. In this case also, $a_x \in \mathcal{A}_{x_1}(v) \cap \mathcal{A}_{y_1}(v)$.

In all cases, we have $\mathrm{d}(x_1, y_1) \leq \mathrm{d}(x, y)$, since the only changes occur at w which is already a disagreement. Hence $\beta(\mathcal{M}_{flip}, \mathcal{C}) \leq 1$ for the modified coupling $\mathcal{C}$. We also have $a_x \in \mathcal{A}_{x_1}(v) \cap \mathcal{A}_{y_1}(v)$ with probability at least κ/nk. Conditionally on this event, we have $x_2(v) = y_2(v) = a_x$ with probability $1/nk$. Thus $\mathrm{d}(x_2, y_2) \leq \mathrm{d}(x, y) - 1$ with probability at least $\kappa(nk)^{-2}$, giving $\sigma_2^2(\mathcal{M}_{flip}, \mathcal{C}) \geq \kappa(nk)^{-2}$.

Hence, as in Theorem 7, $\mathcal{M}_{flip}$ mixes in time

$$\tau(\varepsilon) \leq 2 \left\lceil e\kappa^{-1}k^2n^4 \right\rceil \lceil \log \varepsilon^{-1} \rceil = O\Big(\Delta^2 n^4 \log(\varepsilon^{-1})\Big),$$

which may be compared with (1).

Remark 15. We may apply the comparison of [16] to bound the mixing time of $\mathcal{M}_{Gl}$. This yields a mixing time bound for $\mathcal{M}_{Gl}$ of $\tau(\varepsilon) = O\Big(\Delta^2 n^5 \log(\varepsilon^{-1})\Big)$. This compares with the mixing time bound $O\Big(k^{1+\Delta/6} n^{4+\Delta/6} \log(k) \log(\varepsilon^{-1})\Big)$ stated in [16].

Proof of Proposition 14. In order to prove items (i)-(iii) we must look at Vigoda's coupling in more detail. The coupling is defined in terms of the coupling of flips of clusters, not in terms of coupling choices of vertex of colour. Let σ and τ be two colourings that differ in a single vertex v. Let

$$\begin{aligned}\Gamma_c = &\ \{w : \sigma(w) = c, w \text{ is a neighbour of } v\},\\ \mathcal{D}_{\sigma,c} = &\ \{S_\sigma(v,c)\} \cup \{S_\sigma(w,\tau(v)) : w \in \Gamma_c\}.\end{aligned}$$

$\mathcal{D}_{\tau,c}$ is defined analogously. Further let $\mathcal{D}_\sigma = \cup_{c\in\mathcal{K}}\mathcal{D}_{\sigma,c}$. Note that any cluster of σ not in $\mathcal{D}_\sigma$ occurs also in τ and is flipped with the same probability. The coupling of two chains in states σ and τ is the identity on any flip of a cluster not in $\mathcal{D}_\sigma$. For each $c \in \mathcal{K}$, the flips of clusters in $\mathcal{D}_{\sigma,c}$ and $\mathcal{D}_{\tau,c}$ are coupled together as follows. If $\Gamma_c = \emptyset$ then $\mathcal{D}_{\sigma,c} = \{v\} = \mathcal{D}_{\tau,c}$, and these flips are coupled. If $\Gamma_c \neq \emptyset$, let $w_0 \in \Gamma_c$ and $w_1 \in \Gamma_c$ maximise $|S_\sigma(w,\tau(v))|$ and $|S_\tau(w,\sigma(v))|$ respectively.

- With probability $p_{|S_\tau(v,c)|}/nk$ flip clusters $S_\tau(v,c)$ and $S_\sigma(w_0,\tau(v))$.
- With probability $p_{|S_\sigma(v,c)|}/nk$ flip clusters $S_\sigma(v,c)$ and $S_\tau(w_1,\sigma(v))$.
- For each $w \in \Gamma_c$, couple flips of clusters $S_\sigma(w,\tau(v))$ and $S_\tau(w,\sigma(v))$ maximally over the remaining probability of each flip.
- Any uncoupled probability of flipping these clusters is coupled with the null transition.

This completes the definition of the coupling. Note that with probability at least $1/nk$ both chains make a null transition (giving (i)), for example this occurs if σ chooses vertex $u \neq v$ and colour $\sigma(u)$.

Also note that for each $w \in \Gamma_c$ the flip of $S_\sigma(w,\tau(v))$ is coupled, in some proportion (possibly zero), with each of $S_\tau(v,c)$, $S_\tau(w,\sigma(v))$ and the null transition, and that $S_\tau(v,c) = S_\tau(w,\tau(v))$ has at least two elements.

Now let x and y be any two colourings coupled using this path coupling. Observe that (ii) holds, since if x chooses vertex v and colour $c \in \mathcal{A}_x(v)\cap\mathcal{A}_y(v)$, then at all colourings z in the path x to y we have $\mathcal{D}_{z,c} = S_z(v,c) = \{v\}$.

Suppose that x chooses vertex v and colour $c \in \mathcal{A}_x(v)\setminus\mathcal{A}_y(v)$. Then $S_\sigma(v,c) = \{v\}$. We will prove by induction on the path length that (iii) holds. The base case (path length 0) is trivial. Suppose now that (iii) holds for paths of length $\mathrm{d}(x,y)-1$, and let z be the colouring preceding y in the path. Then either (a) with probability at least κ/nk chain x flips cluster $S_x(v,c) = \{v\}$ and z flips a cluster $S_z(v,c') = \{v\}$, or (b) with probability at least κ/nk chain x flips cluster $S_x(v,c) = \{v\}$ and the chain z makes a null transition.

In case (b), since $\kappa/nk < 1/nk$ we may compose the coupling of x flipping cluster $S_x(v,c) = \{v\}$ and z making a null transition with the coupling of the null transition in both z and y. Hence with probability at least κ/nk chain x

flips cluster $S_x(v,c) = \{v\}$ and the chain y makes a null transition.

For case (a), observe that in the coupling between z and y the flip of cluster $S_z(v,c')$ is coupled either using the identity or in some proportion with flipping $S_y(v,c'')$, $S_y(v,c''')$ and the null transition, where $|S_y(v,c'')| > 1$. If, in addition, $|S_y(v,c''')| > 1$ then since $S_z(v,c')$ is flipped with probability $1/nk$, whereas $S_y(v,c'')$, $S_y(v,c''')$ are each flipped with probability at most p_2/nk, it must be that with probability at least $(1-2p_2)/nk = \kappa/nk$, z flips $S_z(v,c')$ while y makes a null transition. We may therefore compose the couplings of x,z and z,y such that with probability at least κ/nk chain x flips cluster $S_x(v,c) = \{v\}$ and the chain y makes a null transition. If, on the other hand, $|S_y(v,c''')| = 1$, then since $S_y(v,c'')$ is still flipped with probability at most p_2/nk we have probability at least $(1-p_2)/nk$ that z flips $S_z(v,c')$ while y either flips $S_y(v,c''')$ or makes a null transition. By symmetry $S_y(v,c''')$ has remaining probability at least $(1-p_2)/nk$, and since the two flip moves are maximally coupled, z flips $S_z(v,c')$ while y flips $S_y(v,c''')$ with probability at least $(1-p_2)/nk > \kappa/nk$. Thus we may compose the couplings such that with probability at least κ/nk chain x flips cluster $S_x(v,c) = \{v\}$ and the chain y flips $S_y(v,c''') = \{v\}$. This completes the proof. □

References

[1] D. Aldous, Random walks of finite groups and rapidly mixing Markov chains, in *Séminaire de Probabilités XVII 1981/82*, Springer Verlag, Berlin, 1983, pp. 243–297.

[2] R. Bubley, *Randomized algorithms: approximation, generation and counting*, Springer-Verlag, London, 2001.

[3] R. Bubley and M. Dyer, Path coupling: A technique for proving rapid mixing in Markov chains, in *Proc. 38th Annual IEEE Symposium on Foundations of Computer Science*, 1997, pp. 223–231.

[4] P. Creed, Sampling Eulerian orientations of the triangle, preprint.

[5] W. Doeblin, Exposé de la théorie des chaînes simples constantes de Markoff à un nombre fini d'états, Revue Mathématique de l'Union Interbalkanique, 2:77–105, 1938.

[6] M. Dyer and C. Greenhill, A more rapidly mixing Markov chain for graph colourings, *Random Structures and Algorithms* **13** (1998), 285–317.

[7] M. Dyer and C. Greenhill, Random walks on combinatorial objects, in *Surveys in Combinatorics 1999* (J. D. Lamb and D A. Preece, eds.), London Mathematical Society Lecture Note Series **267**, Cambridge University Press, Cambridge, 1999, pp. 101–136.

[8] M. Dyer and C. Greenhill, On Markov chains for independent sets, *Journal of Algorithms* **35** (2000), 17–49.

[9] J. Fehrenbach and L. Rüschendorf, A Markov chain algorithm for Eulerian orientations of planar triangular graphs, in *Mathematics and Computer Science III Algorithms, Trees, Combinatorics and Probabilities* (M. Drmota, P. Flajolet, D. Gardy and B. Gittenberger, eds.), Birkhäuser, 2004, pp. 429–440.

[10] J. Fehrenbach and L. Rüschendorf, Markov chain algorithms for Eulerian orientations and 3-colourings of 2-dimensional Cartesian grids, *Statistics and Decisions* **22** (2004), 109–130.

[11] L. Goldberg, R. Martin and M. Paterson, Random sampling of 3-colourings in $\mathbb{Z}^2$, *Random Structures and Algorithms* **24** (2004), 279–302.

[12] M. Jerrum, A very simple algorithm for estimating the number of k-colorings of a low-degree graph, *Random Structures and Algorithms* **7** (1995), 157–165.

[13] M. Jerrum, *Counting, Sampling and Integrating: Algorithms and Complexity*, Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 2003.

[14] M. Luby, D. Randall and A. J. Sinclair, Markov chains for planar lattice structures, *SIAM Journal on Computing* **31** (2001), 167–192.

[15] M. Mitzenmacher and E. Upfal, *Probability and Computing : Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, Cambridge, 2005.

[16] E. Vigoda, Improved bounds for sampling colorings, *Journal of Mathematical Physics* **41** (1999), 1555–1569.